Project 1: Auxiliary feature learning for small dataset regularization (Supervised learning)

January 24, 2019

1 Description

We have recently seen a wealth of methods that aim at learning from small datasets.

The goal of this project is to see if one can use generic auxiliary tasks to learn features that help regularise the network. The most common of these tasks is to learn an autoencoder and then use the features provided by the autoencoder to learn a discriminative classifier. The idea here is that there is information contained in the features.

2 Tasks

1. Select three different datasets of your choice - this can be from UCB, Kaggle, etc. Load and inspect them.
2. For each of them, train an autoencoder on a small sample of the dataset.
3. Use the learned features to train a standard discriminative neural network - see here for an example.
4. Compare the results (e.g. AUC scores) starting from a small sample data size and progressively moving to the whole dataset.
5. Compare your results with other state-of-the-art methods on the selected datasets.

3 References

1. Data Augmentation: Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D. "mixup: Beyond Empirical Risk Minimization.".
2. Auxiliary Features for RL: Jaderberg, Max, et al. "Reinforcement learning with unsupervised auxiliary tasks."
3. Multi-task learning: Liebel, Lukas, and Marco Körner. "Auxiliary Tasks in Multi-task Learning."
4. Example Transfer Learning: Choi, Keunwoo, et al. "Transfer learning for music classification and regression tasks."
5. Building autoencoders in Keras

4 Dataset examples

1. https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones
2. https://archive.ics.uci.edu/ml/datasets/Audit+Data
3. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra